

Classification of Human Skin Diseases using Data Mining

Qusay Kanaan Kadhim

Department of Computer Science, Al Yarmouk University College, Iraq

Abstract— Many Information can be extracted from the data that are hidden in images. The extraction process can be done using data mining techniques. In this paper, the researcher will use a system based on the decision tree for mining and processing image data. This system will be used for classification of human skin diseases. The researcher will try to use the decision tree and digital image processing principals to detect skin diseases using some features found in a digital image for a skin. The major steps involved in the system are: pre-processing, features extraction and decision tree classifier. This system enhances the classification process to be more accurate. The physicians can make use of this accurate decision tree classification phase for classifying the skin images. The system is designed and implemented on MATLAB and is tested with the images of various databases.

Keywords—Expert system, Data Mining, Image Processing, Skin Diseases, Decision Tree.

I. INTRODUCTION

Knowledge Discovery and Data Mining are rapidly evolving areas of research that are at the intersection of several disciplines, including statistics, databases, AI, visualization, and high-performance and parallel computing. People in business, science, medicine, academia, and government collect such data sets, and several commercial packages now offer general purpose Knowledge Discovery and Data Mining tools. An important Knowledge Discovery and Data Mining goal is to “turn data into knowledge.” For example, knowledge acquired through such methods on a medical database could be published in a medical journal. Knowledge acquired from analyzing a financial or marketing database could revise business practice and influence a management school’s curriculum [1]. Valuable information can be hidden in images, however, few researches discuss data mining on them[2]. Further development of a computer-assisted diagnosis is associated with the use of new intelligent capabilities such as data mining, which allow discovering the relevant knowledge for image analysis and diagnosis from the database of image descriptions. The application of data mining will help to get some additional knowledge about

specific features of different classes and the way in which they are expressed in the image. The decision tree method has been used to classify the medical images for diagnosis. An additional advantage of data mining application for the decision of medical or other tasks is on the long-run the opportunity for creation of fully automatic image-diagnosis systems that could be very important and useful in the case of lack of knowledge for decision-making . Classification is one of the most common applications for data mining[3]. Classification is an important form of knowledge extraction, and can help make key decision[4]. It corresponds to a task that occurs frequently in everyday life. For example, a hospital may want to classify medical patients into those who are at high, medium or low risk of acquiring a certain illness, an opinion polling company may wish to classify people interviewed into those who are likely to vote for each of a number of political parties or are undecided, or we may wish to classify a student project as distinction, merit, pass or fail. Decision Tree have been a powerful and attractive tool in the field of classification, mainly because they produce easily interpretable and organized results. In general computationally efficient and capable of dealing with noisy data[4]. Image mining is more than just an expansion of data mining to image domain. It is an interdisciplinary challenge that draws upon proficiency in computer vision, digital image processing, image extraction, data mining, machine learning, databases, and artificial intelligence[5].

II. BACKGROUND

One of the problems that arises in any collection of data is the classification. The classification is important in many fields. One of these fields is image classification (medical image classification). The classification process depends on the principal of data mining, especially image mining, and can be done using one technique in data mining that is called decision tree. The decision tree will be used to classify medical skin diseases images. This classification can done using a system consists of the following phases on images: pre-processing, features extraction and decision tree classifier. This system is very important and useful for physicians to detect skin diseases

of the human and so to determine the suitable medicine for that disease.

In this paper, the researcher will propose a system to detect and recognize skin diseases in human. The system will distinguish between normal skin and infected skin. The distinction is based on using data mining techniques, specifically the decision tree, and image processing to extract the important features used in classification process. This system can be used by physicians to make the recognition process for skin diseases more accurate depending on a database of skin images.

III. RELATED WORKS

At the existing conditions of computerized skin diagnosis systems, there are some workarounds. Access closed which are still under and research developments. And it is determined some restrictions and barriers in those and therefore this solution tries to overcome the problems that exist together with different take.

a) An automated system for recognizing disease conditions of human skin

In this model, the condition of the skin disease is identified by evaluating skin disease images by using grey normalized symmetrical simultaneous occurrence stencils (GLCM) method. The proposed system is used in an efficient and economical for the automatic recognition of skin diseases. This system is useful for the skin to reduce the error with medical diagnosis. Another is the first test for patients in rural areas, where the good doctors are missing. The system works with relational databases to the storage of implying the need for textual skin images. This system can also work for same type of images directly over feature vectors [6].

b) Image-based diagnosis method

This system mainly focuses on diagnosing diseases of skin that are occurred by viruses and bacteria. This system used image of the diverse area and those images are taken and then machine learning techniques and image processing applied to train the computer to diagnose the skin disease. This is an optional diagnosis method for these skin diseases and it is safe and no risks, side effects or inconveniences from the patient perspective. It also gives advantageous to doctors because it is fast and can be implemented in various ways (mobile phones, computers and digital cameras). And also it can be safely used by non-Specialized medical personnel. First, the patients were clinically analyzed by a professional (dermatologist/medical doctor), then laboratory tests were conducted to foresee and confirm the skin disorder. The doctor then apprehended some

images from the patients whose results showed that they had a viral or bacterial infection [7].

c) Expert System for Diagnosis of Skin Diseases

This system is developed for diagnosing skin diseases which allow user to identify diseases of the human skin to provide advises or medical treatments in a very short time period. The system uses technologies such as image processing and data mining for the diagnosis of the disease of the skin. The image of skin disease is taken and it must be subjected to various processing for noise eliminating and enhancement of image. This image is immediately segmentation of images using threshold values. Finally data mining techniques are used to identify the skin disease and to suggest medical treatments or advice for users[8].

d) The Development of Online Children Skin Diseases Diagnosis System

A system enables the user recognize skin diseases confronted of children through the Internet and make user for advice or Treatments in the shortest period of time. The is based on law and the ahead was used a sequence heuristics engine for development From the system. With this system, to assist and allows the user to Recognition of Pediatric Dermatology through the Internet and offer helpful Proposal the user[9].

IV. ARCHITECTURE OF THE PROPOSED SOLUTION

An Expert System for the diagnosis of skin disease. In our situation we need resolved teams of the pattern first step in styles having a system to retrieve all the images to a particular level as styles is much more obvious to use no noisy and unwanted data then we extract specific features like the colors characteristics can be used to create a model classification the area. With this system model at last classification a predictable disease of the a new image from the skin disease. Building on once more on such diseases prophesied system would ask a user constitute and is based on the system and the answer is decides to the type of disease is the case with used again in the data mining technique. Eventually refers or a medical the tips that are based on the expected to as a result of a skin disease therapy system.

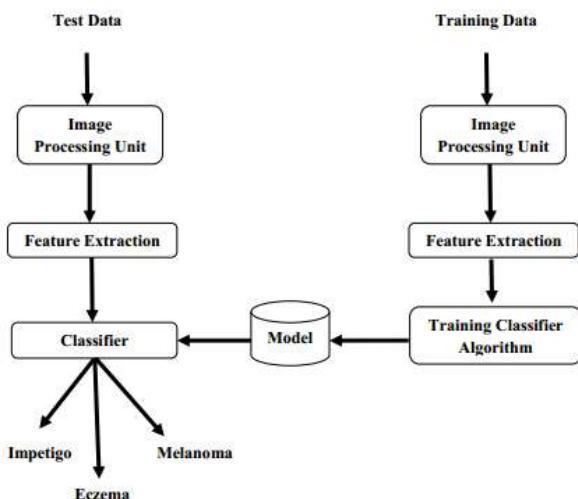


Fig. 1: Architecture of the System

This paper there is plenty of searching in an image data preparing and mining of find out what are most effective method of more accurately using the techniques and getting the best solutions for skin diseases to identify and diagnose. Search consists of three portions .

1. Processed image, and dividing and an advantage essence.
2. Model of classification and attribution skin disease.
3. The medical treatment or advice suggestions.

V. THE EXPERT SYSTEM

A. IMAGE PROCESSING

The image processing is the main part in the design process at an expert system. In the beginning needed to define the affected the area of skin disease that part of The image processing of this process must be It has been implemented. Imaging provide techniques and processes in creating images of human body or samples for clinical purposes medical science or for knowledge discovery. Digital image processing involves the screening of a region for processing and saving this region to a location (possibly a file) for processing [10]. Image process of adjusting digital image so that the result are more suitable for display or further image analysis for example can remove noise sharpen or brighten of digital images, making easier to identify key feature[11]. The detection of skin disorders and their evaluation is divided into some basic steps[12]. The image processing and image filtering techniques. The mathematical concepts of convolution and the kernel matrix are used to apply filters to signals, to perform functions such as extracting edges and reducing unwanted noise the Sobel operator and Gaussian smoothing filter are implemented in MATLAB to achieve the functions previously mentioned, and are evaluated on test images. The effects of adding Gaussian and 'salt and pepper' noise before filtering are then

presented as an approximation to signals that occur in real applications. Pre-processing images before applying other filters is shown to produce improved results when extracting edges from images with noise[13]. Gaussian filtering g is used to blur images and remove noise and detail. In one dimension, the Gaussian function is:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad (1)$$

Using algorithms in the image cutting to the Background Removal from the image process. We separate the area from the disease. A will then be done with retail the image of the picture. Is feature extraction out of and then send the extracted features to the unit data for diagnosis the extraction shows Figure 2 Image processing.

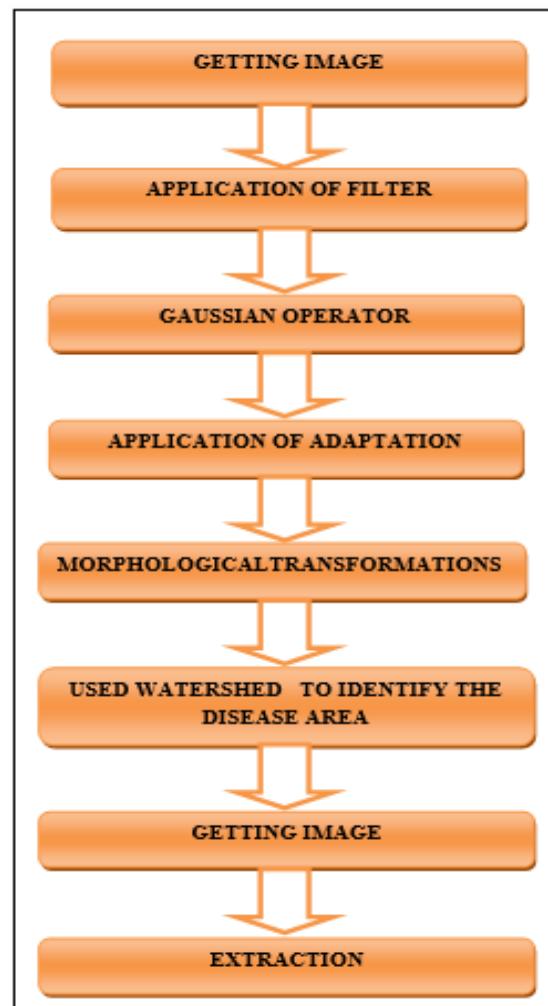


Fig. 2: Image processing

B. DATA MINING

Data mining is the extraction of hidden predictive information and unknown data, patterns, relationships and knowledge by exploring the large data sets which are difficult to find and detect with traditional statistical methods. Data mining it is powerful technology which

will discover most important information from the data warehouse of the organizations. It is a very crucial step that collectively examine large amount of routinely data[14]. Data mining is the process of extracting meaningful information from large database. In Medical field the problem may arise in the era data mining has vital role to predict and diagnosis the disease in early stage with the use of machine learning tool[15]. Data Mining and KDD Process Furthermore before conducting a review and analysis work, we first have to understand what data mining is as the main area of the study declared that data mining came into existence in the middle of 1990's and appeared as a powerful tool that is suitable for fetching previously unknown at tern and useful information from huge dataset. Various studies highlighted that data mining techniques help the data holder to analyze and discover unsuspected relationships among their data which in turn helpful for decisions making stated that a data mining is a technique that deals with the extraction of hidden predictive information from a large database. It uses sophisticated algorithms for the process of sorting through large amounts of data sets and picking out relevant information. Data mining the Analysis step of the Knowledge Discovery in Databases process, or KDD a relatively young and interdisciplinary field of computer science, is the process of extracting Patterns from large data sets by combining methods from statistics and artificial intelligence with database management coined that the term Knowledge Discovery in Databases, or KDD for short, refers to the broad process of finding knowledge in data, and emphasizes the "high-level" application of particular data mining methods. It is of interest to researchers in machine learning, pattern recognition, databases, statistics, artificial intelligence, knowledge acquisition for expert systems and data visualization. Furthermore researcher had presented an outline of the steps of the KDD is the last stage as shown in Figure 3[16].

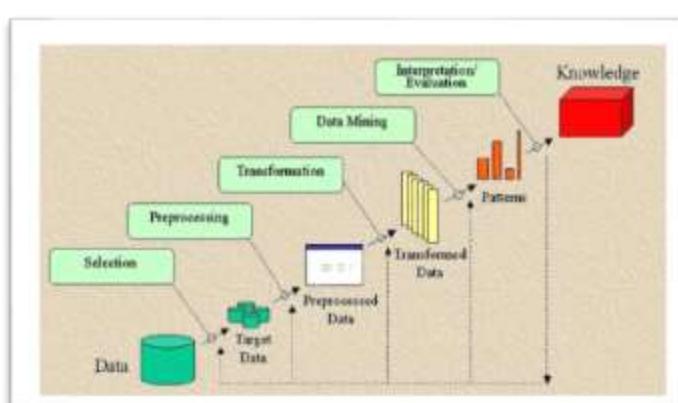


Fig. 3: Stages of Knowledge Discovery Process (KDD)

Data mining algorithms are needed in almost every step in KDD process ranging from domain understanding to knowledge evaluation. It is necessary to identify and evaluate the most common data mining algorithms implemented in modern .Determining performance of data mining solutions require much time and effort. Data mining algorithms may give in better results for one type of problems while others may be suitable for different ones. The need is for algorithms with very high accuracy as medical diagnosis is considered as a significant yet obscure task that needs to be carried out precisely and efficiently[17]. The descriptive data extraction model is to find patterns in the data and identifies the correlation between traits actress in the data. In contrast, the intended to mining predictive input model largely to predict future results.

C. DECISION TREE

Decision tree produces recognition or regression products through a tree construction. It splits a dataset into smaller subsets while at the same time connected decision tree is incrementally developed. The final result is a decision tree with decision nodes and leaf nodes[18]. Decision tree is a predictive data mining techniques often used in clinical medicine to easily visualize, and understand resistant to noise in data. And is applicable in both regression and association data mining tasks[19]. Tree induction decision trees are used to predict and/or classify tree are two phases, the training and implementation. During the training phase, the data set is partitioned iteratively[20]. In this, the target concept is represented in the form a tree, where the tree is built by using the principle of recursive partitioning. In this, attributes are selected as a partitioning attribute or as a node based on the information gain criteria and then the process continues repeatedly for every child node until all attributes are considered and a decision tree is constructed. Some pruning techniques may further be considered so that the size of the tree is reduced and the overfitting is thereby avoided[21]. Be used to create a model for predicting rating assessing skin disease expert system predicted the results is a major task in the system c expert suggests classification algorithms to predict skin disease.

VI. CONCLUSION

The main of this paper is to focus is the use of a proposed program and using image processing can predict and resolve enormous applications. The discovery of knowledge from large amounts of data considering both the result of which is obtained by way of images and the way the questionnaire addressed. That means, we will have a system made up with many of the questions prepared by the system from the user. And the system to

get help answer that was given to the above questions for the skin disease diagnosis. In this system administrator can manage information from a skin disease, symptoms, medical treatment and suggestions and prepared a statement to display the description of the skin disease.

REFERENCES

- [1] Q. Luo, “1. Luo, Q. on Knowledge Discovery and Data Mining WKDD 2008. (2008).on Knowledge Discovery and Data Mining WKDD 2008,” no. January, 2008.
- [2] C. Science, “Image Processing and Image Mining using Decision Trees *,” vol. 1003, pp. 989–1003, 2009.
- [3] P. Smitha, L. Shaji, and M. G. Mini, “A Review of Medical Image Classification Techniques,” *Int. Conf. VLSI, Commun. Instrum.*, no. Icvci, pp. 34–38, 2011.
- [4] S. Nandgaonkar, R. Jagtap, P. Anarase, B. Khadake, and A. Betale, “Image mining of textual images using low-level image features,” in *Proceedings - 2010 3rd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2010*, 2010, vol. 9, pp. 588–592.
- [5] C. L. Devasena, “An Experiential Survey on Image Mining Tools , Techniques and Applications,” vol. 3, no. 3, pp. 1155–1167, 2011.
- [6] N. Yadav, V. Kumar, and N. Director, “Skin Diseases Detection Models using Image Processing: A Survey Utpal shrivastava,” *Int. J. Comput. Appl.*, vol. 137, no. 12, pp. 975–8887, 2016.
- [7] C. L. Aruta, C. R. Calaguas, J. K. Gameng, M. V. Prudentino, A. Anthony, and C. J. Lubaton, “Mobile-based Medical Assistance for Diagnosing Different Types of Skin Diseases Using Case-based Reasoning with Image Processing,” vol. 3, pp. 115–118, 2015.
- [8] A. A. L. C. Amarathunga, E. P. W. C. Ellawala, G. N. Abeysekara, and C. R. J. Amalraj, “Expert System For Diagnosis Of Skin Diseases,” *Int. J. Sci. Technol. Res.*, vol. 4, no. 1, pp. 174–178, 2015.
- [9] M. M. Yusof, R. A. Aziz, and C. S. Fei, “The Development of Online Children Skin Diseases Diagnosis System,” vol. 3, no. 2, pp. 231–234, 2013.
- [10] D. Okuboyejo, O. Olugbara, and S. Odunaike, “Automating Skin Disease Diagnosis Using Image Classification,” *Proc. World Congr. Eng. Comput. Sci.*, vol. II, pp. 23–25, 2013.
- [11] S. Kumar, “Image Processing for Recognition of Skin Diseases,” vol. 149, no. 3, pp. 37–40, 2016.
- [12] H. Motion, “Detection and Evaluation of Skin Disorders By One of Photogrammetric Image Analysis Methods,” vol. XXXIX, no. September, pp. 537–542, 2012.
- [13] S. Kim, “Applications of Convolution in Image Processing with MATLAB,” *Univ. Washingt.*, pp. 1–19, 2013.
- [14] S. Kaur and R. K. Bawa, “Future Trends of Data Mining in Predicting the Various Diseases in Medical Healthcare System,” *Int. J. Energy, Inf. Commun.*, vol. 6, no. 4, pp. 17–34, 2015.
- [15] D. Sindhuja, “A Survey on Classification Techniques in Data Mining for Analyzing Liver Disease Disorder,” vol. 5, no. 5, pp. 483–488, 2016.
- [16] H. Leopold, W. Kipruto Cheruiyot, and S. Kimani, “A Survey and Analysis on Classification and Regression Data Mining Techniques for Diseases Outbreak Prediction in Datasets,” *Int. J. Eng. Sci.*, pp. 2319–1813, 2016.
- [17] P. Mahindrakar and M. Hanumanthappa, “Data Mining In Healthcare : A Survey of Techniques and Algorithms with Its Limitations and Challenges,” *Int. J. Eng. Res. Appl.*, vol. 3, no. 6, pp. 937–941, 2013.
- [18] M. Preethi, “Analyzing Human Skin Texture using Machine Learning Approaches,” *Int. J. Comput. Appl.*, vol. 136, no. 1, pp. 975–8887, 2016.
- [19] K. Danjuma and A. O. Osofisan, “Evaluation of Predictive Data Mining Algorithms in Erythematous Squamous Disease Diagnosis,” no. Cvd, p. 10, 2015.
- [20] T. D. Kale, “Informatics Application of Data Mining Techniques to Discover Cause of Under- Five Children Admission to Pediatric Ward : The Case of Nigist Eleni Mohammed Memorial Zonal Hospital,” vol. 6, no. 1, pp. 1–14, 2015.
- [21] M. Kumarasamy, “An Improving Performance of Data Mining Using Intelligent Agent System Methods,” vol. 6495, no. 12, pp. 104–110, 2016.